DOCUMENT RESUME

ED 225 742 RC 013 783

AUTHOR Loyd, Brenda H.
TITLE Analysis of Content-Related Item Bias for Anglo and Hispanic Students.
PUB DATE Mar 82
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, March 19-23, 1982).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Anglo Americans; Comparative Analysis; *Content Area Reading; English; *Hispanic Americans; Intellectual Disciplines; *Item Analysis; Mathematics; Natural Sciences; Secondary Education; Social Studies; *Test Bias
IDENTIFIERS *ACT Assessment

ABSTRACT

A random sample of October, 1979, test sites was selected to assess the potential bias of the ACT Assessment. From this sample, sites which tested at least five Anglos and five Hispanics were included; the final sample contained a total of 876 students (403 Anglo and 373 Hispanic). Using the primary hypothesis, "There is no interaction between the items and ethnic groups," a tnree factor (items, sex, and ethnic group) analysis of variance design was used to separately analyze the four tests (English, mathematics, social studies, and natural science). The item by ethnic group interactions were statistically significant for two of the four content tests, English usage and social studies. Items significantly contributing to the item by ethnic group interactions were analyzed further. Within the English usage test, five reading passages (six items) contributed to the item by ethnic group interaction; three items favored Anglos and three favored Hispanics. Seven items were identified for the Social Studies Reading tests, four favored Anglos and three favored Hispanics. Over the two tests, there were seven items favoring Anglos and six favoring Hispanics; therefore, the tests could not reasonably be said to favor one group decisively more than the other. Four tables are appended. (BRR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED225742

Analysis of Content-Related Item Bias for Anglo and Hispanic Students

Brenda H. Loyd
University of Virginia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Brenda H. Loyd

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Paper presented at the Annual Meeting
of the American Educational Research Association
New York
March, 1982

RC013783

Analysis of Content-Related Item Bias for Anglo and Hispanic Students

Brenda H. Loyd
University of Virginia

INTRODUCTION

As an issue in standardized testing, bias has received considerable attention in recent years. In the course of discussing this issue many different definitions of bias have been proposed. One definition which has some generality in a measurement context is that which describes bias as differential validity. A test which is differentially valid is one which measures different constructs for two (or more) groups.

One approach to assessing tests for differential validity would focus on the results of performance on each test as a whole. With the use of external criteria one could determine the extent to which differences in overall test performance for two (or more) groups were accurately reflected in differences in future performance. However, it is also of interest to examine differential validity at the item level. In the absence of specific external criteria, it is still possible to assess the relative nature of group performance on a given test. Bias or differential validity at the item level is determined relative to the entire network of items which compose a given test.

Recent literature is replete with descriptions of methods for detecting item bias (Shepard, Camilli & Averill, 1980; and Rudner, Gelson & Knight, 1980). These investigators have compared techniques for detecting item bias empirically and with Monte Carolo studies. The amount of evidence favoring one technique over the others has been

less than conclusive. However, a procedure employing an analysis of variance design agrees closely with many of the item bias techniques which have been investigated; it also has the advantage of allowing simultaneous comparison of many groups as well as types of groups such as race and sex. This procedure was therefore selected to provide an estimation of item bias in the tests investigated in the present study.

Potential "bias" for this study will be defined as a significant item by group interaction in an analysis of variance procedure. Such a significant interaction will be considered to be an indication of possible differential validity of the items. This procedure allows two groups to perform differently without regarding all differences between groups as bias. It also allows one to identify those items which are differentially easy or difficult for a given group compared to the average difference in difficulty of all times for the two groups. Such items are then considered for possible bias and are subjected to further investigation to try to determine reasons for such differential group performance.

METHOD: Data

The ACT Assessment contains four tests which are developed to measure academic abilities in subject areas traditionally associated with high school curricula and to provide information concerning probable academic accomplishments of the examinees on the college level.

The four test areas are: English, mathematics, social studies, and natural science. The English Usage Test measures understanding and use of the basic elements of correct and effective writing. The Mathematics Usage Test assesses mathematical reasoning and problem-solving ability. The Social Studies Reading Test and the Natural Sciences Reading Test measure the critical reasoning and problem-solving

abilities required in social studies and natural science, respectively.

All four tests employ a multiple choice format, with English Usage containing 75 items,; Mathematics Usage containing 40; Social Studies Reading containing 52; and Natural Sciences Reading containing 52.

Several tests forms are created each year for administration to a population of examineees across the nation. The test form selected for this investigation was administered nationally during the 1979-80 school year.

From all test sites which administered this form in October, 1979, a random sample of sites was selected. From this random sample, sites which tested at least 5 anglo and 5 hispanic examinees (ethnic group indicated by self report) were included in the study. The records of equal numbers of anglo and hispanic students from these sites were then used in the analysis. Sex of the examinees was also a factor of interest. The final sample contained a total of 876 students; 403 anglo and 373 hispanic.

Procedure

On all students' records, item responses for each of the four tests were coded "1" for a correct response, and "0" for an incorrect response or an omit. For each item, the item difficulty, average item score, for each subgroup (grouped by ethnic group and sex designation) was determined.

A three factor analysis of variance design was used to analyze the data. Each of the four tests was analyzed separately. The factors were items, sex and ethnic group. The primary hypothesis of interest was: "There is no interaction between the items and ethnic groups."

The analysis of variance procedure was used to test this hypothesis for each of the four tests.

In order to stabilize the variance of the item difficulties, an arcsin transformation was used. The transformation of each item difficulty (p) for each sex group of each ethnic group was of the form $\Psi = 2 \arcsin \sqrt{p}$. This made the variance of each cell $1/n_{jk}$ where n_{jk} was the number of students in sex group j and ethnic group k.

For each of the 4 tests, the experimental design can be represented as in Figure 1.

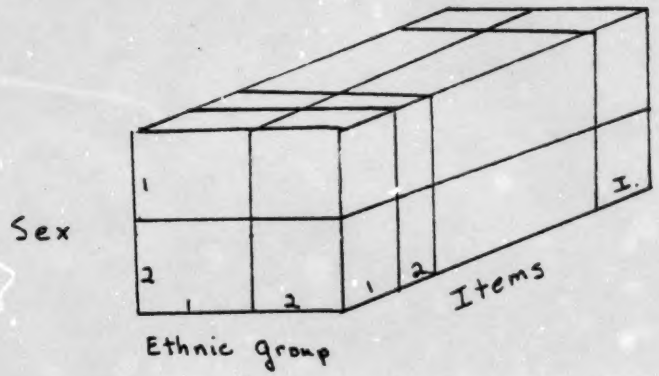

Figure 1

When the interaction terms were found significant, the interaction contrast for each item was obtained. The contrast formed for a significant item by ethnic group interaction is as follows:

$$\psi_{ik} = \phi_{ik} - \bar{\phi}_i - \bar{\phi}_k + \bar{\bar{\phi}}$$

where

ϕ_{ik} = the average arcsin transformed item difficulty of the i^{th} item for the k^{th} ethnic group across the two sex groups

$\bar{\phi}_i$ = the mean of the arcsin transformed item difficulties for the i^{th} item across the two sex groups and the two ethnic groups

$\bar{\phi}_k$ = the mean of the arcsin transformed item difficulties for ethnic group k across items and sex groups

$\bar{\bar{\phi}}$ = the overall mean.

A significant contrast identified a particular item as one which contributed significantly to the particular interaction. Any item so designated was examined in an attempt to ascertain a plausible reason for the items interaction characteristics. The number and type of item by content and skills classification were also examined.

Results

The summary tables for the analysis of variance for each of the four tests are presented in Tables 1. The item by ethnic group interactions were statistically significant ($p < .05$) for two of the four content tests. These were the English usage and social studies tests. The item by ethnic group interactions for the mathematics and natural sciences tests were not statistically significant.

Interaction contrasts were computed for each item by ethnic group in the English usage and social studies tests. Interaction contrasts were analyzed separately for each of the two tests. A summary

of the number of items contributed significantly to the item by ethnic group interaction is presented in Table 2. Also presented are the total number of items included in each analysis and the number of identified items which favor each ethnic group.

The items which were identified as contributing significantly to either item by group interaction are presented by test and description in Table 3. Included with each item are the content classification, the group favored by the item, and the item difficulties for the designated groups. For purposes of clarity in description and understanding, the item difficulties rather than the arcsin transformed item difficulties are presented.

While the main effects of ethnic groups were not the primary interest in the investigation, the average item difficulty for each group for each test is presented in Table 4.

Discussion

The items significantly contributing to the item by ethnic group interactions were analyzed further for consistency in content, skills classification and item difficulty. The items of the English Usage tests all have references to specific reading passages. Seven reading passages compose the English Usage test for this form of the test. Of these seven passages, items identified as contributing to the item by ethnic group interaction were based on five reading passages. Of the two items identified for the "Reading" topic, one favored anglos and one favored hispanics; of the four additional identified items each in a different passage, two favored anglos and two favored hispanics. There is no substantial evidence that the passage content definitively favors either group.

A review of the skills classification of the six identified items of the English Usage test reveals no consistency of classification for the three items which favor each group. Although a variety of specific skills are represented, there is some evidence that adjectives and adverbs presented in a series may be more difficult for hispanic students and correct use of prepositions and pronouns may be easier than expected. It may be worthwhile to see if this trend is found in other forms of this test or in other tests.

The item difficulties of these seven identified items showed no specific relationship to the particular group favored.

For the Social Studies Reading test, seven items were identified, four favoring anglos and three favoring hispanics. The social studies test consists of items which are directly related to reading passages and of items which are non-passage related and therefore depend more upon background knowledge to answer the questions. Four reading passages are included in the exam and present information in the four areas of history, government, economics, and social sciences (psychology, sociology, etc.) Items contributing significantly to the item by group interaction were identified for the economics and government passages. Of the two items identified for the economics passage, one favored anglos and one favored hispanics. Of the three items identified for the government passage, two favored hispanics and one favored anglos. No significant trend was indicated.

Of the 15 non-passage related items, two were identified and both favored anglos. One item was about psychology and the other economics. There is some indication that items which depend strictly on background information are significantly more difficult for hispanics. This may indicate the need for further investigation to determine if high

school curricula in any general sense is different for the two groups. However on the basis of two items, no definitive conclusion can be reached.

The significant item by group interactions in the English Usage test and the Social Studies Reading test may indicate that the content and nature of these tests are more sensitive to interaction with ethnic group because of their dependence on language. Over the two tests, there were seven items favoring anglos and six favoring hispanics. Therefore the test could not reasonably be said to favor one group decisively more than the other. There are however some trends which may deserve further examination in order to substantiate the findings suggested by this analysis.

References

Cardall, Carolyn and Coffman, W. E. A method for comparing the performance of different groups on the items in a test. Research Bulletin 64-67. Princeton, N.J.: Educational Testing Service, 1964.

Cleary, T. A. and Hilton, T. L. An investigation of item bias. Educational and Psychological Measurement, 1968, 28, 61-75.

Plake, B. S. and Hoover, H. D. The comparability of equal raw scores obtained from in-level and out-of-level testing: One source of discrepancy between in-level and out-of-level grade equivalent scores. Journal of Educational Measurement, 1979, 16, 271-278.

Rudner, L. M., Gelson, P. R. and Knight, D. L. A Monte Carlo comparison of seven biased item detection techniques. Journal of Educational Measurement, 1980, 17, 1-10.

Shepard, L., Camilli, G. and Averill, M. Comparison of sex procedures for detecting test item bias using both internal and external ability criteria. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, April 1980.

TABLE 1

English Usage Test:

Analysis of Variance Summary Table

Source	df	MS	F
Between			
Sex	1	.315	
Ethnic Group	1	4.015	
Sex by Ethnic Group	1	.086	
Within			
Items	74	.336	
Items by Sex	74	.008	
Items by Ethnic Group	74	.009	3.25*
Residual	74	.003	

*p .05

Mathematics Usage Test:

Analysis of Variance Summary Table

Source	df	MS	F
Between			
Sex	1	.893	
Ethnic Group	1	2.328	
Sex by Ethnic Group	1	.211	
Within			
Items	39	.1976	
Items by Sex	39	.004	
Items by Ethnic Group	39	.003	1.29
Residual	39	.002	

*p .05

TABLE 1 (cont.)

Social Studies Reading Test:

Analysis of Variance Summary Tables

Source	df	MS	F
Between			
Sex	1	.613	
Ethnic Group	1	1.832	
Sex by Ethnic Group	1	.091	
Within			
Items	51	.521	
Items by Sex	51	.004	
Items by Ethnic Group	51	.005	2.21*
Residual	51	.002	

*p .05

Natural Sciences Reading Test:

Analysis of Variance Summary Table

Source	df	MS	F
Between			
Sex	1	.741	
Ethnic Group	1	1.926	
Sex by Ethnic Group	1	.052	
Within			
Items	51	.314	
Items by Sex	51	.002	
Items by Ethnic Group	51	.001	.82
Residual	51	.002	

*p .05

TABLE 2

Summary of Items Significantly Contributing

to the Items by Ethnic Group Interaction

ENGLISH USAGE	
# with significant Contrasts	6
Total # in Test	75
# Favoring Hispanics	3
# Favoring Anglos	3
SOCIAL STUDIES READING	
# with Significant Contrasts	7
Total # in Test	52
# Favoring Hispanics	3
# Favoring Anglos	4

TABLE 3

Items Identified by Interaction Contrasts

ENGLISH USAGE

Item	Passage # and Topic	Description of Skill	Group Favored	Item Difficulty A	Item Difficulty H
A	2 Spring	logic of adjective/ adverb placement	A	.84	.61
B	3 Reading	punctuation of adjective/ adverb series	A	.69	.47
C	3 Reading	use of prepositions	H	.64	.66
D	5 Rocky Mountains	logic of metaphor	A	.76	.58
E	6 Carver	use of pronouns	H	.77	.77
F	7 History of Language	punctuation of adjective series	H	.33	.30

TABLE 3 (cont.)

SOCIAL STUDIES READING

Items	Content	Group Favored	Item Difficulty A	Item Difficulty H
A	Economics/passage related item	A	.79	.62
B	Economics/passage related item	H	.42	.42
C	Government/passage related item	H	.75	.75
D	Government/passage related item	A	.70	.58
E	Government/passage related item	H	.62	.65
F	Psychology/non-passage related	A	.69	.49
G	Economics/non-passage related	A	.56	.37

TABLE 4

Average Item Difficulties

Test	Ethnic Group H	Ethnic Group A
English Usage	.52	.64
Mathematics Usage	.39	.50
Social Studies Reading	.43	.52
Natural Sciences Reading	.39	.49